

Analisis Penerapan Sistem untuk Rekomendasi Bidang Olimpiade Siswa SMPN 2 Cikupa Berbasis Web Menggunakan *Naive Bayes*

Ruskanda Danda Anugrah^{*1}, Zurnan Alfian, S.Kom., M.Kom.²

^{1,2,3}Universitas Pamulang; JL. Raya Puspitek No. 46, Buaran, Kec. Serpong, Kota Tangerang Selatan
Banten 15310, (021) 7412566

^{1,2,3}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

e-mail: ^{*1}ruskandadandaanugrah@gmail.com, ²dosen02678@unpam.ac.id

Abstrak

Sistem pemilihan siswa untuk olimpiade di SMPN 2 Cikupa masih bersifat manual dan subjektif berdasarkan pengamatan guru. Penelitian ini bertujuan mengembangkan sistem rekomendasi bidang olimpiade berbasis web menggunakan algoritma Gaussian Naive Bayes untuk memberikan rekomendasi objektif berdasarkan data akademik siswa. Metode penelitian menggunakan pendekatan kuantitatif dengan pengumpulan data nilai rapor 855 siswa kelas 7 dan 8 yang mencakup bidang IPA, IPS, dan Matematika. Sistem dikembangkan menggunakan Python dengan framework Flask, template engine Jinja2, dan database MySQL. Hasil pengujian menunjukkan akurasi model sebesar 79,53% dengan precision tertinggi pada bidang MTK (87,34%) dan recall tertinggi pada bidang MTK (85,19%). Pengujian Black Box menunjukkan seluruh fungsi sistem berjalan sesuai harapan. Sistem berbasis web yang dikembangkan memudahkan akses siswa dan pihak sekolah dalam memperoleh rekomendasi bidang olimpiade secara objektif dan terstruktur, menggantikan pendekatan konvensional yang subjektif. Berbeda dengan penelitian sebelumnya yang fokus pada klasifikasi jurusan di tingkat SMA, penelitian ini secara spesifik mengarahkan siswa SMP pada bidang olimpiade berdasarkan analisis nilai rapor secara otomatis dan real-time.

Kata kunci— Naive Bayes, Sistem Rekomendasi, Olimpiade, Machine Learning, Website

I. PENDAHULUAN

Pendidikan merupakan aspek fundamental dalam membentuk sumber daya manusia berkualitas. Dalam konteks pendidikan di Indonesia, peningkatan mutu peserta didik menjadi prioritas utama, tidak hanya dalam aspek akademik reguler tetapi juga pengembangan potensi melalui ajang kompetitif seperti Olimpiade Sains Nasional (OSN). OSN berperan penting dalam membentuk karakter kompetitif dan ketangguhan akademik siswa (Ginting et al., 2024).

Menurut teori Differentiated Model of Giftedness and Talent (DMGT) yang dikembangkan François Gagné, potensi alami siswa perlu diarahkan melalui proses yang melibatkan berbagai faktor termasuk motivasi dan lingkungan (Barbier et al., 2023). Bakat siswa dapat berkembang menjadi prestasi nyata jika didukung faktor intrapersonal seperti motivasi dan

lingkungan yang mendukung seperti guru serta akses sumber belajar.

Namun, proses pemilihan bidang olimpiade di SMPN 2 Cikupa masih dilakukan secara subjektif berdasarkan pengamatan guru atau permintaan siswa, yang belum tentu sesuai dengan potensi sebenarnya. Penelitian menunjukkan nilai rapor siswa, terutama mata pelajaran tertentu seperti matematika, menjadi indikator utama memprediksi keberhasilan akademik (Gunawan et al., 2024). SMPN 2 Cikupa belum memiliki sistem berbasis data yang dapat memberikan rekomendasi objektif dan efisien.

Perkembangan teknologi informasi khususnya *machine learning* memberikan solusi untuk permasalahan ini. Algoritma Naive Bayes yang bekerja berdasarkan prinsip probabilitas dapat mengklasifikasikan data siswa secara spesifik dan efektif (Prairi et al., 2025). Penelitian sebelumnya menunjukkan penerapan Naive Bayes dalam klasifikasi

data pendidikan memberikan rekomendasi objektif berdasarkan data tersedia (Khairy et al., 2024).

Penelitian ini mengembangkan sistem rekomendasi bidang olimpiade berbasis web menggunakan algoritma *Gaussian Naive Bayes* di SMPN 2 Cikupa. Berbeda dengan penelitian sebelumnya yang fokus pada klasifikasi jurusan di tingkat SMA (Perkasa & Purwiantono, 2023; Yudha & Putri, 2022), penelitian ini secara spesifik mengarahkan siswa SMP pada bidang olimpiade (IPA, IPS, atau Matematika) secara otomatis berdasarkan analisis nilai rapor real-time. Sistem dirancang untuk menganalisis data nilai rapor siswa dan memberikan rekomendasi bidang olimpiade secara objektif, membantu guru dalam pengambilan keputusan dan memberikan arah tepat bagi siswa mengembangkan potensinya.

II. METODE PELAKSANAAN

Penelitian ini menggunakan metode pengumpulan data sekunder berupa nilai rapor dari 855 siswa kelas 7 dan 8 SMPN 2 Cikupa yang terbagi dalam tiga bidang, yaitu IPA dengan 256 siswa, IPS dengan 405 siswa, dan MTK dengan 194 siswa. Proses pengumpulan data dilakukan melalui wawancara dengan kepala sekolah dan observasi terhadap sistem yang sedang berjalan di sekolah tersebut.

Sebelum data digunakan dalam penelitian, dilakukan tahap pra pemrosesan yang meliputi proyeksi fitur linear dan normalisasi data. Proyeksi fitur linear dipilih untuk menggabungkan mata pelajaran yang memiliki keterkaitan konseptual kuat, seperti agregasi nilai IPA dan MTK untuk fitur rata-rata IPA, nilai PKN dan IPS untuk fitur rata-rata IPS, serta nilai Matematika dan IPA untuk fitur rata-rata MTK. Tujuan normalisasi menggunakan Min-Max Scaling dengan rentang 1-100 adalah untuk menyetarakan skala data dan meningkatkan performa model klasifikasi.

Pengembangan sistem dilakukan menggunakan Python dengan framework Flask untuk backend dan web service, library Scikit-learn untuk implementasi *Gaussian Naive Bayes*, serta HTML, CSS, JavaScript dengan Bootstrap untuk frontend. Database MySQL digunakan untuk penyimpanan data. *Gaussian Naive Bayes* dipilih karena efisiensi komputasinya yang tinggi dan kemampuannya bekerja optimal dengan data berukuran kecil hingga menengah, sesuai dengan kondisi data sekolah. Data dibagi menjadi 80% data latih (684 sampel) dan 20% data uji (171 sampel).

Evaluasi model dilakukan menggunakan tiga metrik utama untuk mengukur performa sistem. Metrik pertama adalah akurasi yang mengukur proporsi

prediksi yang benar dari keseluruhan prediksi. Metrik kedua adalah precision yang menghitung proporsi prediksi positif yang benar dari seluruh prediksi positif yang dibuat. Metrik ketiga adalah recall yang mengukur proporsi kasus positif yang berhasil teridentifikasi dengan benar dari seluruh kasus positif yang ada.

Untuk memastikan sistem berfungsi dengan baik, dilakukan pengujian menggunakan metode Black Box Testing yang bertujuan untuk memverifikasi bahwa seluruh fungsionalitas sistem dapat berjalan sesuai dengan kebutuhan pengguna tanpa melihat struktur internal kode program.

III. HASIL DAN PEMBAHASAN

Berdasarkan pengujian yang dilakukan pada 171 sampel data uji, model *Gaussian Naive Bayes* menghasilkan performa yang cukup baik. Hasil evaluasi model dapat dilihat pada tabel berikut:

Tabel 1. Hasil Evaluasi Model

Metrik	Nilai
Akurasi	79,53%
Precision IPA	81,63%
Precision IPS	62,79%
Precision MTK	87,34%
Recall IPA	78,43%
Recall IPS	69,23%
Recall MTK	85,19%

Model berhasil memprediksi dengan benar 136 dari 171 sampel yang diuji, dengan rincian 40 sampel bidang IPA, 27 sampel bidang IPS, dan 69 sampel bidang MTK. Performa terbaik dicapai pada kategori MTK dengan precision sebesar 87,34% dan recall sebesar 85,19%, yang menunjukkan kemampuan model dalam mengidentifikasi 85,19% peserta didik yang sebenarnya termasuk dalam kategori MTK.

Dalam komparasi dengan algoritma lain, hasil perbandingan kinerja dapat dilihat pada tabel berikut:

Tabel 2. Komparasi Kinerja Algoritma

Algoritma	Akurasi	Precision IPA	Recall IPA
Gaussian Naive Bayes	79,53%	81,63%	78,43%
SVM	85,38%	92,86%	76,47%
Decision Tree	90,64%	90,57%	94,12%
Random Forest	93,57%	94,12%	94,12%

Meskipun Random Forest menunjukkan performa yang lebih superior dengan akurasi mencapai 93,57%, *Gaussian Naive Bayes* tetap dipilih sebagai algoritma

utama dalam penelitian ini karena memiliki keunggulan dalam hal kecepatan komputasi dan efisiensi memori yang sangat ideal untuk sistem rekomendasi yang bekerja secara real-time.

Sistem berhasil diimplementasikan dengan berbagai fitur lengkap yang disesuaikan dengan kebutuhan setiap jenis pengguna. Untuk siswa, sistem menyediakan fitur registrasi, login, input nilai rapor, melihat hasil rekomendasi, riwayat tes, profil siswa, dan pengisian kuisioner feedback. Bagi admin, sistem dilengkapi dengan dashboard monitoring, pengelolaan data siswa, laporan analisis hasil rekomendasi, dan pengaturan akun. Sedangkan untuk super admin, sistem menyediakan fitur manajemen akun pengguna, pengelolaan data siswa dan laporan, pengaturan bobot rekomendasi, manajemen kuisioner feedback, serta pengaturan metode dan dataset.

Pengujian Black Box yang dilakukan pada seluruh fitur sistem menunjukkan hasil yang valid untuk semua skenario pengujian. Proses registrasi dan login berjalan dengan lancar tanpa hambatan. Input nilai rapor tersimpan dengan akurat dalam sistem. Rekomendasi bidang olimpiade yang diberikan sesuai dengan data akademik siswa. Fitur-fitur yang tersedia untuk admin dan super admin berfungsi dengan optimal. Selain itu, proses perubahan password berhasil dilakukan tanpa mengalami error sama sekali.

Hasil penelitian ini menunjukkan bahwa sistem rekomendasi berbasis Gaussian Naive Bayes dapat memberikan rekomendasi yang objektif berdasarkan data nilai rapor siswa. Akurasi sebesar 79,53% menunjukkan performa yang baik dan sejalan dengan penelitian sebelumnya yang menggunakan Naive Bayes untuk klasifikasi jurusan dengan akurasi 87,85%. Performa model yang tertinggi pada bidang MTK dengan precision 87,34% dan recall 85,19% mengindikasikan bahwa karakteristik nilai matematika dan IPA lebih distingtif atau memiliki pembeda yang lebih jelas dibandingkan dengan bidang IPS. Hal ini didukung oleh penelitian terdahulu yang menunjukkan bahwa nilai matematika merupakan prediktor yang kuat untuk keberhasilan akademik siswa.

Hasil ini menunjukkan bahwa Gaussian Naive Bayes cukup efektif digunakan dalam lingkungan sekolah dengan data terbatas karena efisiensi komputasinya yang tinggi dan tidak memerlukan tuning parameter yang kompleks. Secara praktis, sistem ini membantu guru mengidentifikasi potensi siswa secara lebih objektif dan mempercepat proses seleksi peserta olimpiade, yang sebelumnya memakan waktu berhari-hari menjadi hanya beberapa detik. Bagi siswa, sistem memberikan panduan yang lebih jelas mengenai bidang olimpiade yang sesuai dengan kemampuan

akademiknya, sehingga dapat memfokuskan persiapan dengan lebih terarah.

Keterbatasan performa yang terjadi pada bidang IPS dengan precision 62,79% dan recall 69,23% kemungkinan disebabkan oleh ketidakseimbangan dalam distribusi data, di mana terdapat 405 sampel IPS yang mencapai 47,4% dari total data, sementara sampel IPA hanya 29,9% dan MTK 22,7%. Penelitian di masa mendatang dapat mengatasi masalah ini melalui penerapan teknik oversampling atau undersampling untuk menyeimbangkan distribusi data. Sistem berbasis web yang dikembangkan dalam penelitian ini memberikan kemudahan akses dan transparansi dalam proses pemberian rekomendasi, sehingga dapat menggantikan pendekatan subjektif yang selama ini digunakan secara konvensional. Fitur-fitur lengkap yang tersedia untuk siswa, admin, dan super admin memastikan bahwa pengelolaan data dapat dilakukan secara sistematis dan terstruktur dengan baik.

IV. SIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi bidang olimpiade berbasis web menggunakan algoritma Gaussian Naive Bayes di SMPN 2 Cikupa dengan akurasi 79,53%. Model menunjukkan performa terbaik pada bidang MTK (Precision 87,34%, Recall 85,19%) dan dapat memberikan rekomendasi objektif berdasarkan data nilai rapor siswa. Sistem berbasis web yang dikembangkan memudahkan akses siswa dan sekolah dalam memperoleh rekomendasi bidang olimpiade secara objektif, menggantikan pendekatan konvensional yang subjektif.

Keterbatasan penelitian ini terletak pada ketidakseimbangan distribusi data antar bidang yang menyebabkan performa bidang IPS lebih rendah. Penelitian selanjutnya dapat menggunakan teknik balancing data seperti SMOTE (Synthetic Minority Over-sampling Technique) atau undersampling untuk meningkatkan performa bidang IPS. Selain itu, penambahan fitur non-akademik seperti minat siswa dan riwayat kegiatan ekstrakurikuler dapat meningkatkan akurasi rekomendasi secara lebih komprehensif.

DAFTAR PUSTAKA

- Barbier, K., Struyf, E., Verschueren, K., dan Donche, V. 2023. Fostering Cognitive and Affective-Motivational Learning Outcomes for High-Ability Students in Mixed-Ability Elementary Classrooms: A Systematic Review.

European Journal of Psychology of Education. 38(1):83-107.

Ginting, F.W., Syafrizal, M., Mellyzar, Rejeki, N.S., dan Harahap, E. 2024. Bimbingan Olimpiade Sains Nasional Bidang Fisika dan Kebumian Bagi Siswa SMA/MA. *Jurnal Pengabdian Masyarakat*. 5(2):120-128.

Grinberg, M. 2018. *Flask Web Development*, Ed.2. O'Reilly Media, Sebastopol.

Gunawan, A., Tine, M., dan Murwaningtyas, C. 2024. Prediksi Keberhasilan Akademik Menggunakan Nilai Rapor. *Jurnal Pendidikan Indonesia*. 12(3):215-228.

Khairy, D., Alharbi, N., Amasha, M.A., Areed, M.F., Alkhalaf, S., dan Abougalala, R.A. 2024. Prediction of Student Exam Performance Using Data Mining Classification Algorithms. *Education and Information Technologies*. 29(4):4235-4258.

Kuhn, M. dan Johnson, K. 2019. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press, Boca Raton.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., dan Brucher, M. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12:2825-2830.

Perkasa, K.B.P.Y. dan Purwiantono, F.E. 2023. Sistem Rekomendasi Jurusan Menggunakan Algoritma Naive Bayes Gaussian Berbasis Web. *J-Intech (Journal of Information and Technology)*. 11(2):361-370.

Prairi, M.Z., Alfian, Z., dan Harefa, K. 2025. Kecocokan Keputusan Pohon Algoritma pada Kimia Organik: Perbandingan ROC AUC Keputusan Pohon dan Ketetangaan. *Journal of Innovative and Creativity*. 5(2):10689-10700.

Yudha, M. dan Putri, D.I. 2022. Pemanfaatan Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI. *Jurnal Tekno Kompak*. 16(2):176-185.